Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 1996 Proceedings

Americas Conference on Information Systems (AMCIS)

8-16-1996

Base Rate Expectations and the Detection of Errors in Data: Direct Experience versus Information

Barbara D. Klein
School of Management, University of Michigan-Dearborn, bdklein@fob-fl.umd.umich.edu

Base Rate Expectations and the Detection of Errors in Data: Direct Experience versus Information

Barbara D. Klein
Assistant Professor
bdklein@fob-f1.umd.umich.edu
School of Management
University of Michigan-Dearborn
4901 Evergreen Road - 133 FOB
Dearborn, MI 48128

Background

There is strong evidence (e.g., Laudon, 1986; Morey, 1982) that data stored in organizational databases are neither entirely accurate nor complete. If undetected, errors in data may significantly affect business outcomes. Two main approaches to this problem are (1) validating data as they are input to or stored in databases (e.g., Morey, 1982) and (2) depending on users to detect and correct errors.

A research program examining the efficacy of the second approach is underway. To date, studies completed in this program have shown that users in several business professions (actuarial science, consumer product management, inventory management, and municipal bond analysis) detect errors in data and that error detection goals and incentive structures affect error detection performance (Klein, 1995). However, two previous studies in this research program have shown mixed results for the effect of expectations about the base rate of errors in data (Klein, 1995). In one study, subjects had different expectations about the base rate of errors stemming from direct experience. In the other, expectations about the base rate of errors were manipulated through information in a controlled experimental setting. This mixed finding motivates the specific research question addressed in the study described here. The present study employs a laboratory experiment to examine the effect of base rate expectations developed through direct experience with data containing errors.

The methodology and results of the two prior studies showing mixed results for the expectations about the base rate of errors factor are briefly summarized below.

1. Experience-Based Expectations. Field interviews were conducted with five professionals working in three domains (consumer product management, inventory management, and municipal bond analysis). Domains were selected to vary along the dimension of the base rate of errors. Base rate expectations of the subjects interviewed in this study had been developed through several years of direct experience working with data. A relationship between expectations about the base rate of errors and self-reports of error detection was found ($p<.05$).

2. Information-Based Expectations. A laboratory experiment was conducted to examine the impact of base rate expectations on performance in the detection of data errors. Subjects in the experiment assumed the role of consultants performing pension calculations for client firms. The expectation construct was operationalized through a memo providing information about the error rate. There were three versions of this memo. In the high base rate condition, the memo stated that the clients for whom the subjects would be performing the pension calculations tend to provide data containing a lot of errors. In the low base rate condition, the memo stated that the clients for whom the subjects would be performing the pension calculations tend to provide data that is mostly free of errors. No mention of the accuracy of the data provided by the clients in the past was made in the memo read by subjects in a control condition. The actual number of errors was constant across all conditions in this experiment. No main effect of the base rate expectations factor was found. Thus, we were unable to conclude that the performance of subjects with expectations of a high base rate of errors is better than the performance of subjects with expectations of a low base rate of errors.

The conclusion drawn from these studies is that users will respond to the experience of detecting errors in data rather than to information about the base rate of errors. It appears that expectations about the error rate are most strongly influenced by hands-on experience with error detection. However, the observed relationship between experience-based expectations and error detection performance is based on self-reported incidents of error detection collected through field interviews. The present study seeks to bolster our confidence about this relationship by studying the relationship between experience-based expectations about the base rate of errors and objective measures of error detection performance in a controlled laboratory setting.

A research model has been developed using signal detection theory (Green and Swets, 1966), Campbell's (1990) theory of individual task performance, theories of adaptive decision making (Payne, 1982), and theories of expertise (Ericsson and Chase, 1982).

Objectives and Propositions

The goal of this research is to improve our understanding of the conditions under which individuals detect errors in data. Two propositions are tested in the study.

Proposition 1: Experience-based expectations about the base rate of errors in data influence performance in error detection. This proposition is based on signal detection theory (Green and Swets, 1966) and theories of expertise (Ericsson and Chase, 1982).

Proposition 2: There is an interaction between experience-based expectations about the base rate of errors in data and incentive structures. This proposition is based on signal detection theory (Green and Swets, 1966), Campbell's (1990) theory of individual performance, and theories of adaptive decision making (Payne, 1982). The specific prediction is that increasingly powerful levels of incentives will be associated with greater differences in error detection performance among different levels of base rate expectations.

Design and Methodology

The study is being conducted as a laboratory experiment. The experimental task is from the domain of employee benefits. The task and the experimental materials have been successfully used in a prior study within this research stream (Klein, 1995). Subjects use personnel data to calculate the pension benefit for the employees of two firms. Errors were embedded in the data based on an analysis of the task in consultation with a domain expert. Subjects have ninety minutes to perform the experimental task.

There are two independent variables: 1. expectations about the base rate of errors (high, low, or control) and 2. incentive structure ($100 prize awarded based on the number of errors detected, $100 prize awarded randomly, no prize). The expectations about the base rate of errors factor will be manipulated through direct experience rather than through information about the base rate of errors as was done in the prior study. In the high base rate condition, subjects will perform the pension calculations using a dataset containing a high error rate (30 percent of the records contain an error) before performing the pension calculations for a second target company. In the low base rate condition, subjects will perform the pension calculations using a dataset containing a low error rate (10 percent of the records contain an error) before performing the pension calculations for a second target company. For the control condition, subjects will only perform the pension calculations for the target company. A filler task will proceed the pension calculations for the control group to control for subject fatigue. The levels of the incentive structure used in this study have been found to affect the detection of data errors in prior research (Klein, 1995).

Dependent variables are the proportion of errors detected and the proportion of accurate values incorrectly identified as errors. Signal detection theory also suggests a measure of discriminability (the ability to distinguish accurate from inaccurate data) which will be compared across experimental conditions. Only performance on the (second) target company will be measured.

162 subjects recruited from upper-level undergraduate and graduate-level business courses will participate in the experiment. Students are used because, compared to experienced professionals, they have weaker *a priori* expectations about the base rate of errors in data.

A questionnaire will be used to check whether subjects' perceptions are consistent with the levels of the treatments to which they were exposed. Four questions (two for the expectations of the base rate factor and two for the incentive structure factor) will be included.

The effects of the expectations about the base rate of errors factor and the incentive structure factor will be investigated using standard ANOVA models. If interaction or main effects are found, the Tukey method of multiple comparisons will be used to test for differences in performance between specific treatment groups.

Conclusion

A synthesis of the results of this investigation and the results of the prior studies in this research stream will improve our understanding of the precise relationship between expectations about the base rate of errors in data and error detection performance. The finding that base rate expectations developed through direct experience do not affect error detection will provide a strong challenge to the validity of the relationship between base rate expectations and error detection previously found in the field interview study. This outcome would suggest that base rate expectations and error detection performance are unrelated. The alternative finding that expectations developed through direct experience affect performance will provide strong support for the conclusion that users respond to the experience of detecting errors in data rather than to information about the error rate.

An understanding of the conditions under which users do and do not detect errors in data has implications for the design and management of information systems. The finding that error detection can be improved by base rate expectations developed through direct experience is an important prerequisite to the development of prescriptive interventions to improve error detection. While it may not be easy to provide users with such experiences, an issue for future investigation is whether users might benefit from learning about specific errors that their colleagues have found. Such knowledge might boost their perceptions about the base rate of errors and lead to performance improvements.

References

Campbell, J.P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M.D. Dunnette and L.M. Hough (Eds.), **Handbook of Industrial and Organizational Psychology** (2nd ed., Vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologists Press, Inc.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. **American Scientist**, 70, 607-614.

Green, D.M., & Swets, J.A. (1966). **Signal detection theory and psychophysics**. New York: Wiley.

Klein, B. D. (1995). Base rates and payoffs in the detection of errors in data**. Proceedings of the First Americas Conference on Information Systems**, Pittsburgh, Pennsylvania, August 25-27, 1995, pp. 381-383.

Laudon, K.C. (1986). Data quality and due process in large interorganizational record systems. **Communications of the ACM**, **29**, 4-11.

Morey, R. C. (1982). Estimating and improving the quality of information in a MIS. **Communications of the ACM**, **25**, 337-342.

Payne, J.W. (1982). Contingent decision behavior. **Psychological Bulletin**, 92, 382-402.